DOCUMENT RESUME

ED 073 448 CS 000 430

AUTHOR	Ruddell, Robert B., Ed.
TITLE	Accountability and Reading Instruction: Critical Issues.
INSTITUTION	National Council of Teachers of English, Urbana, Ill.
PUB DATE	73
NOTE	51p.
AVAILABLE FROM	National Council of Teachers of English, 1111 Kenyon Road, Urbana, Ill. 61801 (Stock No. 00037, $1.50 non-member, $1.35 member)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Behavior; *Behavioral Objectives; Behavior Patterns; Curriculum; Curriculum Development; *Educational Accountability; *Performance Contracts; Performance Factors; Reading; *Reading Instruction; *Reading Tests

ABSTRACT

The intent of the Commission on Reading of the National Council of Teachers of English in developing this group of articles was to provide insight into critical issues related to accountability and reading instruction. The initial presentation by James Laffey develops a brief historical analysis of educational accountability, followed by a discussion of the alternatives and problems that must be understood in developing an accountability system. Richard Hodges identifies the nature and sources of behavioral objectives as he examines assumptions related to goals of instruction, measurement, and the content and methodology of learning. Kenneth Goodman's paper on testing concerns reading tests, design problems in constructing reading tests, and abuses of tests. Advantages as well as limitations and disadvantages of performance contracting are discussed by Mary Galvan. Galvan concludes her discussion by identifying needs and making recommendations useful in developing curriculum-oriented guidelines for schools interested in performance contracting. The concluding discussion by Harold Herber identifies a range of critical issues based on the preceding discussions. (Author/WR)

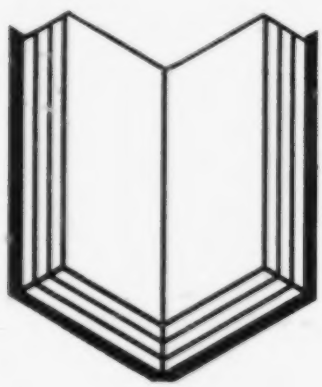
ED 073448

ACCOUNTABILITY AND READING INSTRUCTION

CRITICAL ISSUES

ROBERT B. RUDDELL, EDITOR
UNIVERSITY OF CALIFORNIA, BERKELEY

FILMED FROM BEST AVAILABLE COPY



U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

CS 000 430

NATIONAL COUNCIL OF TEACHERS OF ENGLISH
1111 KENYON ROAD, URBANA, ILLINOIS 61801

PERMISSION TO REPRODUCE THIS COPY RIGHTED MATERIAL HAS BEEN GRANTED BY
Nat'l Council of Teachers of English
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

NATIONAL COUNCIL OF TEACHERS OF ENGLISH

COMMISSION ON READING
Robert B. Ruddell, University of California, Berkeley, *Director*
Helen G. Bacon, University of California, Davis
Charlotte K. Brooks, District of Columbia Public Schools
Charles R. Cooper, State University of New York at Buffalo
Margaret Early, Syracuse University
Thomas J. Edwards, State University of New York at Buffalo
Carolyn Wicker Field, Free Library of Philadelphia, Pennsylvania
Mary Galvan, Texas Education Agency
Kenneth S. Goodman, Wayne State University
Doris V. Gunderson, U.S. Office of Education
Harold L. Herber, Syracuse University
Richard E. Hodges, University of Chicago
William A. Jenkins, Portland State University
James L. Laffey, Indiana University
Kenneth W. Lund, Scott, Foresman and Company
Janet Schubb, University of California Laboratory School, Berkeley
Richard L. Venesky, University of Wisconsin
Robert A. Bennett, San Diego Unified School District; ex officio
Robert F. Hogan, NCTE Executive Secretary; ex officio

EDITORIAL BOARD Richard Corbin, Charlotte S. Huck, Owen Thomas, Richard Lloyd-Jones, Roy C. O'Donnell, Robert F. Hogan, Paul O'Dea
STAFF DESIGNER Norma Phillips Meyers

ISBN 0-8141-0003-7
Library of Congress Catalog Card Number 73-79000
NCTE Stock Number 00037

Copyright © 1973 by the National Council of Teachers of English
All rights reserved. Printed in the United States of America

FOREWORD

The education profession is presently confronted with an unparalleled demand to account for student achievement and teacher performance. Systems and management models from space age technology are being applied directly in the formulation and evaluation of reading and language curricula. The "scientific management" offered by a range of publishers through instructional systems, management systems, and performance-based contracts creates at least the impression that pedagogy is rapidly being transformed into educational technology. Many state legislatures have completed, or are in the process of completing, legislation which will "insure" that adequate evaluation of the educational product will take place. Our "accountability era" of the 1970s is in many ways similar to education's "efficiency era" of the early 1900s. Teachers and administrators are intensely concerned about key issues related to instructional objectives, testing, cost accounting, and about the way in which these concepts will affect instruction and more importantly the learner.

From the time of the creation of the Commission on Reading by the National Council of Teachers of English in 1970, its membership has been intensely concerned about issues related to accountability and reading instruction. This is reflected in Commission-sponsored meetings at the November, 1971 Las Vegas NCTE Convention and the May, 1972 Detroit IRA Convention. The discussions that follow derive from these meetings and pro-

vide a basis of understanding while raising critical issues related to accountability and reading instruction.

The initial presentation by James Laffey develops a brief historical analysis of educational accountability followed by a discussion of the alternatives and the crucial problems which must be understood in developing an accountability system. Three issues introduced in Laffey's paper—behavioral objectives, testing, and performance contracting—are discussed in depth by Richard Hodges, Kenneth Goodman, and Mary Galvan.

Richard Hodges identifies the nature and sources of behavioral objectives as he examines underlying assumptions related to goals of instruction, measurement, and the content and methodology of learning. He then considers the consequences that derive logically from the assumptions. His critical analysis of the formulation of objectives provides important cautions and guidelines for reading instruction.

The penetrating paper on testing by Kenneth Goodman concerns reading tests, statistical fallacies in reading tests, design problems in constructing reading tests, and abuses of tests. Goodman poses critical questions test-makers must consider if their instruments are to become valid and move beyond sophisticated test theory. Finally, he speculates on the nature of future reading tests, giving special consideration to the evaluation of reading achievement "as it really occurs in natural language."

Mary Galvan's discussion provides an overview of recent developments in performance contracting, ranging from concerns with the early Texarkana project to insights derived from her performance-contracting work in the Texas Education Agency. Advantages as well as severe limitations and disadvantages of performance contracting are discussed. The consideration given to legal questions deserves careful attention. Galvan concludes her discussion by identifying needs and recommendations useful in developing curriculum oriented guidelines for schools interested in performance contracting.

The brief concluding discussion by Harold Herber identifies a range of critical issues based on the preceding discussions. Although the intent of these summary issues relates to reading instruction, the reader will quickly recognize their general applicability. Herber's call for a re-examination of our priorities in light

of the process of accountability deserves careful thought and consideration by all members of the profession.

The intent of the Commission in developing this discussion was to provide insight into critical issues related to accountability and reading instruction. Hopefully, these papers will prove to be of decision-making value as educators at all levels carefully examine the vast array of theoretical and applied problems that must be considered in an "accountability era."

Robert B. Ruddell, Director
Commission on Reading
National Council of Teachers of English

TABLE OF CONTENTS

Foreword iii

ACCOUNTABILITY: A BRIEF HISTORY AND ANALYSIS 1
James L. Laffey, *Indiana University*

SOME ASSUMPTIONS ABOUT BEHAVIORAL OBJECTIVES AS RELATED TO READING 13
Richard E. Hodges, *University of Chicago*

TESTING IN READING: A GENERAL CRITIQUE 21
Kenneth S. Goodman, *Wayne State University*

PERFORMANCE CONTRACTING: A ONCE-OVER 35
Mary Galvan, *Texas Education Agency*

ACCOUNTABILITY: A SUMMARY STATEMENT 45
Harold L. Herber, *Syracuse University*

ACCOUNTABILITY:
A BRIEF HISTORY AND ANALYSIS

An article by Stanley Elam, "The Age of Accountability Dawns in Texarkana," leaves the reader with the distinct impressions that (1) accountability is a commonly understood concept and (2) it is a relatively new educational concept.[1] Both ideas are erroneous.

The purpose of this paper is to trace the roots of modern concepts of accountability to an earlier period in educational history known as the "efficiency era," to discuss the origins of more recent ideas of accountability, and to analyze some problems relevant to both versions of accountability.

Early Origins of Accountability

The "efficiency era" in education began around 1900 and [illegible] about 1925. It was an age when scientific management [illegible]ed itself as the panacea for solving all the problems of the schools. Even though scientific management failed in this endeavor, as many educators would have predicted, the years and experiences did point out one of the hard realities educators have to face. Schools exist in a cultural context, and often the cultural context dictates how the schools operate.

[1] Stanley Elam, "The Age of Accountability Dawns in Texarkana," *Phi Delta Kappan* 51 (June 1970), 509-14.

The cultural context of the "efficiency era" was that of business and industry. Raymond E. Callahan states that "the rise of business and industry to a position of prestige and influence [resulted in] America's subsequent saturation with business and industrial values and practices." [2] And as business and industry's policies and leaders began to exert themselves as a major cultural influence, it became apparent that educators and school administrators were in extremely vulnerable positions. Within this cultural setting, the "efficiency expert" entered the arena to save the schools from their own inefficiencies.

In 1911, Frederick Taylor, an industrial engineer, began to expound a system of scientific management. In fact, because of America's concern for efficiency, Taylor became prominent nationally. This in turn led Taylor to pronounce that his principles had *universal applicability;* for he said "his principles could be applied with equal force to all social activities: to the management of homes; the management of our farms; the management of the business of our tradesmen." [3]

Due to Taylor's influence, and the influence of others like him, the remaining years of the second decade of the twentieth century were devoted to criticizing the schools for their inefficiencies and asking why the schools of America were not as efficient as business and industrial organizations. One layman clarified the issue by stating that "if they [the schools] were as efficient as business and industry, then they could provide the public with results that could readily be seen and measured." [4] The response by educators to these critical observations led to the exploratory development of many standardized evaluation forms and tests.

Possibly one of the most important single events during this time was the hiring of a school administrator named Frank Spaulding. Spaulding began his career as superintendent of schools in Newton, Massachusetts—a "burial ground for superintendents." The school board in Newton was notorious for its concern with spending. In other words, the dollar was the educa-

[2] Raymond E. Callahan, *Education and the Cult of Efficiency* (Chicago: University of Chicago Press, 1962), p. 5.

[3] *Ibid.*, p. 43.

[4] *Ibid.*, p. 48.

tional criterion in Newton. And it is possible to suggest that with the hiring of Frank Spaulding by the Newton schools, educational accountability had its start.

For the school year 1911-1912, Spaulding initiated a cost accounting system in the Newton schools. He reported to a group of school administrators and discussed not only per pupil expenditure for the school year but also per unit costs of specific subjects and percentage of the total budgets invested in specific subject areas. In fact, his analysis led Callahan to the following conclusions:

> Spaulding's conception of scientific management obviously amounted to an analysis of the budget. By a study of *local considerations* he meant a study of the per-pupil costs and pupil-recitation costs. His scientific determination of *educational value* turned out to be a determination of *dollar value*. His decisions on what should be taught were made *not on educational, but on financial grounds*. This was not the first time nor was it to be the last. But this occasion was particularly unfortunate because it was presented to leading administrators from all over the nation by one of their leaders and because it clothed this business philosophy and practice with the mantle of science.[5]

While this event relates to the origin of modern concepts of accountability, other significant events occurring during that era reflect an entire range of activities, influencing school accountability not only then but now.

Year	*Event*	*Related to accountability of*
1911	Educational cost accounting recommended	School administration
1912	Briefing of school superintendents on use of school survey	School administration
1912	Measurement of educational efficiency (adaptation of F. Taylor method to education)	Student-Teacher
1915	Student efficiency test	Student-Teacher
1916	Report of the Gary or Platoon Plan—more efficient use of school space	School administration

[5] *Ibid.*, p. 73.

1917	Teacher efficiency record	School administration-Teacher
1923	Increase in class size to decrease school expenses recommended	School administration
1952	Advent of school public relations	School administration

Each event influenced a different aspect of school operation: 1911—budget or financial record keeping; 1912—school surveys as a technique for reporting school effectiveness; 1912 and 1913—precise measurement as a method for determining student learning; 1916—more effective use of school space; 1917—early attempts to assess teacher effectiveness; 1923—manipulation of class size to decrease per pupil costs; and 1925—administrative reporting to influence community support for schools.

Recent Origins of Accountability

While it is not always possible to relate acts of individuals directly to a set of given results, recent events concerned with education and the schools suggest that our political and educational leaders are primarily responsible for the renewed interest in accountability. A number of spokesmen and leaders in government agencies have recommended that accountability and performance contracts be awarded to contractors willing to negotiate such contracts. Lessinger and Allen state that educators should be "required to describe and measure the behavior expected of each student upon completion of the program they propose for funding."[6] This *Phi Delta Kappan* article is one product of earlier work done by Lessinger who at one time was affiliated with the U.S. Office of Education. Much of Lessinger's work was prepared for the specific purpose of renewing or developing ideas relating to accountability in education.

Political leaders have been influential also in stimulating interest in modern concepts of accountability. In his 1969 "State of Education" address, President Richard M. Nixon stated that

[6] Leon M. Lessinger and D. M. Allen, "Performance Prosposals for Educational Funding: A New Approach to Federal Resource Allocation," *Phi Delta Kappan* 51 (November 1969), 136-37.

there is "a corresponding need in the school systems of the nation . . . to begin responsible open measurement of how well the educational process is working. It matters very little how much a school building costs; it matters a great deal how much a child in that building learns. . . ." He continues, "We have, as a nation, too long avoided thinking of the productivity of schools." He also pointed out that although we are spending more on education in this country than in the entire rest of the world (65 billion dollars), we are not getting a significant return on that investment.

Although President Nixon did mention that we spend 65 billion dollars in education, he failed to mention that we spend less than two per cent of our national budget on education. Secondly, he failed to mention that the high cost of education in the United States is related to the high standard of living in this affluent society. So, although there were 65 billions spent on education, relatively speaking (i.e., relative to the gross national income) this is a very small sum. The most serious error in the President's message probably was the statement, "We have, as a nation, too long avoided thinking of the productivity of the schools." The history of American education in the twentieth century is filled with the concern of educators for the "productivity of the schools."[7]

Another possible reason for the renewed interest in accountability is that it is viewed once more by educators as a way to answer critics of the schools. If any institution has been society's scapegoat, it has been the school. Schools have been and are extremely vulnerable to public criticism. Since they are supported and controlled locally, public criticism and pressure can be exerted in various ways.

One impact on the curriculum has been to re-introduce scientific management techniques as a means of eliminating waste and of improving the efficiency of the schools. Consequently, a number of program evaluation techniques (e.g., PERT, CIPP) were promoted as unique efforts in scientific management of schools and school-related organizations. An analysis of the models for efficient management promoted in the past two

[7] Callahan, *Education and the Cult of Efficiency*.

decades and of the scientific management techniques introduced into the schools in the 1920s reveals some parallels. According to Callahan, the essentials of the scientific management method applied in the 1920s were (1) the measurement and comparison of results; (2) the analysis and comparison of the conditions under which given results are secured—especially the means and time employed in given results; and (3) the consistent adoption and use of those means that justify themselves most fully by their results, abandoning those that fail to so justify themselves.[8]

An examination of these three principles of the scientific method reveals the kernel thoughts of current accountability concepts and also of scientific management techniques, i.e., (1) measurement—which first requires a definition of the objectives to be measured; (2) analysis of the learning conditions; and (3) modification or adoption of only those means that are efficient or effective.

Also in these kernel thoughts are some ideas presented by Stephen Barro. He discusses five different approaches to accountability and states, "The focus here is on accountability for effective use of resources."[9] Specific proposals include articulation of goals, introduction of output-oriented management methods and —most importantly—regular comprehensive evaluation of new and ongoing programs.

The earlier concepts restricted accountability to the internal operations of the schools. Newer concepts provide for input from external agents. Barro's article also highlights this difference. Essentially, if the schools were to change under the earlier concepts, teacher or principal behavior within the schools had to change. Newer concepts focus on changes that may (1) emphasize internal changes in the school system; (2) create external evaluations or educational audits; (3) initiate performance incentives for school personnel; (4) make it possible through performance contracts for agencies and institutions outside the schools to enter the schools; (5) cause a shift in the decision-making powers from the central administration of large school units to individual school districts or principals; or (6) result in the development of

[8] *Ibid.*, p. 68.

[9] Stephen Barro, "An Approach to Developing Accountability Measures for the Public Schools," *Phi Delta Kappan* 52 (December 1970), 197.

alternative educational systems. These six alternatives seem to reveal the major differences between the older and newer concepts of accountability. Modern accountability concepts are much more comprehensive and provide for many more alternatives.

Problems with Accountability

One accountability problem, which was identified earlier in the century and still remains, is that of measurement. Educators in the earlier part of the century recognized the difficulties of measuring educational achievement. One comment made in 1913 is appropriate today: "If scientific measurement is to be accomplished, we must have units or scales of measurement which will enable us to make measurements which are verifiable by other observers. We may not hope to achieve progress except as such measuring sticks are available or may be derived." [10] What is apparent in this statement is that educators realized the necessity for valid and reliable measuring instruments even then.

More recently, there has been a growing dissatisfaction with the technical development of modern standardized tests.[11] This is especially true in reading, where standardized tests are inappropriate for evaluating the reading behavior which should be evaluated in any program holding students or teachers accountable. Earlier educators saw the need for developing appropriate tests. Modern educators, after having evaluated valid and reliable standardized tests, are calling for new kinds of tests. Glaser and Nitko suggest that new kinds of tests be developed to measure instructional outcomes. "Tests which are used for making instructional decisions demand special characteristics—characteristics that are different from the mental test model that has been successfully applied in aptitude testing work." [12] They go on to state

[10] Callahan, *Education and the Cult of Efficiency*, p. 101.

[11] See Roger Farr, *Reading: What Can Be Measured* (Newark, Del.: ERIC/CRIER and International Reading Association, 1969); also Stephen D. Klein, "The Uses and Limitations of Standardized Tests in Meeting the Demands for Accountability," *UCLA Evaluation Comment* 2 (January 1970), 1-7.

[12] Robert Glaser and Anthony J. Nitko, "Measurement in Learning and Instruction," *Educational Measurement*, ed. Robert L. Thorndike (Washington, D.C.: American Council on Education, 1971), p. 652.

that special types of criteria need to be developed. Of significance are (1) the creation of items from stated objectives; (2) the creation of interpretive materials for such tests in terms of test content and criteria for performance as well as references to norms for other test-takers; and (3) the extensive application of test performance to domains of content from which the test items were sampled. In essence, modern educators are calling for criterion-referenced tests which interpret an individual's performance with respect to a defined behavioral criterion and which are not limited to a comparison with the performance of other individuals. In addition, there is a need for other newer methods of measuring student behavior related to the affective domain. Finally, there is a need for personnel education that will prevent misadministration, incorrect scoring, and misinterpretation of test results.

The focus for modern evaluation schemes related to accountability is broader than earlier attempts at accountability. Teacher behavior was measured by rating sheets; principal behavior, by rating scales.[13] Little information was gathered on more complex aspects of student, teacher, or administrative behavior. Consequently, early efforts at accountability were less than effective. The instruments used to observe and rate teachers and principals were neither valid nor reliable. Reality has not changed drastically in fifty years; theory has.[14]

A concern not mentioned in the earlier history of accountability but significant to modern education is curriculum articulation. In a number of current accountability systems (such as the Banneker School in Gary, Indiana), curricular articulation was not considered initially. Only after the Indiana State Department of Public Instruction intervened did the contractor attempt to justify and articulate the curriculum already implemented in the school. In the Texarkana Project, a lack of articulation or even of teacher cooperation was noted when some students performed successfully in the reading part of the curriculum (as indicated by teacher grades) but failed in the regular curriculum of the

[13] Callahan, *Education and the Cult of Efficiency.*

[14] N. L. Gage (ed.), *Handbook of Research on Teaching* (Chicago: Rand McNally, 1965).

English classes. Both reading and English were supposed to be part of the same curriculum. More important than the grading practices in the curriculum are the teaching and the learning which should take place from one grade level to the next. It is not apparent that articulation was identified as a problem in earlier accountability programs. In today's program, there is at least the awareness that it may be a problem.

It should be noted, however, that the issue of curriculum articulation in accountability is more serious in some fields than in others. For example, because there is no scientifically based reading curriculum, articulation is less critical in reading than in some other areas. The curriculums of the most widely used instructional approach (basal and basic readers) are developed by authors and publishers. The skill sequence and content are usually determined by what logically and pedagogically appears best. The language used is often not determined on the basis of what is known about children's language usage but by what appears pedagogically wise. Probably the most critical aspect of this issue is whether or not the content of the material is palatable. Will the children read and enjoy the material? Will the material in some way parallel the life experiences of the children? Is the material in line with what we know about children's interest and tastes?

In addition to measurement and articulation, another consideration slighted was the long-term effects of the instructional system. In some current accountability projects, contractors refused to negotiate a delayed testing program to determine whether the student gains were permanent or whether they would disappear after six months. Short-term gains are characteristic in education. It seems essential that in any educational program, attention be given to the long-range effects. Nevertheless, it seems that modern educators are not as concerned with this problem as they should be. Otherwise, accountability contracts would require that the contracting agency be responsible for long-range results as well as short-term results. The contracts would never be for only one year. Only long-term contracts allow the schools to evaluate accountability as a mechanism of change.

Long-range planning relates to a fourth problem: the problem of attaining affective goals in accountability contracts. Former

Commissioner of Education, James Allen, made issue of the affective domain in his Right to Read speech. His main goal for the 1970s included *not only the skill to know how to read but also the desire to read.* From the descriptions of some modern accountability contracts, it is apparent that their contractors have given this some thought. However, only measures of cognitive skills were used to evaluate the Texarkana Project. Methods and techniques focussing on the affective domain were not apparent in earlier attempts to deal with accountability. In more recent times, it seems that educators are knowledgeable about accountability but fail to give due consideration to the assessment of affect—in reading instruction at least. This must change.

An aspect of accountability focussed differently today is teacher impact. From one viewpoint, accountability may require the development of new relationships with the teachers, teachers' unions, and other organizations. From a second viewpoint, accountability may need close observation and control. *Will individual teachers still be able to practice the art of teaching? Or will teaching become skill-oriented drudgery?* Although in earlier times teachers' unions were not a concern, the "art" of teaching was. Today the two concerns should be combined. Teachers' unions will play a role in negotiating accountability contracts; therefore, they should reflect a concern for the art of teaching as well as the skill of teaching.

Involving the community in the educational decision-making process is also a problem in the accountability process. There have been a number of effective innovations lost to the schools because of insensitivity to the local community. Today parents want to participate in any decision-making process that affects the school life of their child. In earlier times, as well as today, the community stimulated the growth of accountability systems. It is not quite clear what role the community will play beyond this in modern accountability programs.

A final and probably the most crucial problem related to accountability is evaluation. At the present time, different views of accountability suggest different approaches to evaluating school programs. Some are concerned with student achievement. Others focus attention on teacher behavior. Still others are concerned with administrative behavior. Whatever the dimensions

of accountability are, the evaluation plan must correspond to the accountability plan, and a sufficient number of trained personnel must be available to implement the evaluation system. In earlier attempts at accountability, comprehensive evaluation plans were unknown. Today they are considered vital to the success of a program.

Summary

In this paper, concepts of accountability were related to an earlier efficiency era in education, described in terms of the more recent origins of today's accountability, and analyzed in regard to problems relevant to both older and newer ideas of accountability. Obviously, accountability needs to be defined operationally to meet the contracts and demands of the governing agency. The realities and failures of accountability in the past will enable educators to avoid those failures in the future.

You may perceive that I have a positive attitude toward accountability. I assure you that I do. But only because I think that accountability can have a positive impact on the process of educating children. And because I believe that the only way accountability as such will have a positive impact is for educators to reject cynical and skeptical attitudes toward accountability and to seize it as an opportunity to hold the public accountable for the resources and support needed by the schools.

SOME ASSUMPTIONS ABOUT BEHAVIORAL OBJECTIVES AS RELATED TO READING

In this short paper, I shall set forth some views about *behavioral objectives,* some principal assumptions that underlie them, some statements of their asserted purposes, and some issues in relation to behavioral objectives as they pertain to reading. Before doing so, however, it is useful to place this brief discussion of behavioral objectives into its larger educational context.

As a social institution, the education enterprise has always been liable to society for its successes and failures in educating the nation's children and youth; "accountability" is not a recent introduction into American public education. Yet, today, during a time when national and international turmoil is causing us to look deeply at our traditional values, even greater demands are being made than has historically been the case for the nation's schools to reaffirm their credentials as society's formal educational agencies. "Performance contracts," "voucher plans," and the like are all symptomatic of pervading social forces which bring pressure upon school systems, their personnel, and their instructional programs to demonstrate both efficiency and effectiveness in educating the young—in short, to be accountable.

The larger (and more significant) question of "education for what?" cannot be examined in the limited space available here. For present purposes, we shall assume that *education may be construed as a change in behavior toward some acceptable goal.*

The set of goals for which education is responsible includes the acquisition of basic understandings of the world in which we live, basic competencies for effective participation in the larger society, and the basic values upon which the society rests. In short, education is the process for the transmission of the premise that certain knowledge, skills, and values are deemed necessary for individual fulfillment and for the maintenance of the social order. It is these educational outcomes that are subject to a "cost accounting" and give particular support to the notion of *behavioral objectives.*

What Are Behavioral Objectives?

Behavioral objectives are claimed by their proponents to be a technical device which can improve the effectiveness of schooling. Tangible evidence that objectives are in fact being reached is provided in the form of *observed* behavior manifested by a student in relation to a particular objective; the *statement* of the objective identifies what behavior a student should demonstrate if the objective is achieved. Here is an example of a behavioral objective in reading as applied to a population of pupils:

> To increase the reading achievement and skills of disadvantaged first-grade students as measured by the total scores obtained on the Stanford Reading Achievement Test, Primary Battery 1, in which the following is obtained: (a) a 25% decrease in the number of students entering the second grade who are one-half or more years behind grade level than was the case for the previous year, and (b) an average of .3 increase in the grade level achievement of students entering the second grade over those of the preceding year.[1]

We should remind ourselves that educational objectives are *value judgments* about the ends of education; thus behavioral objectives also are value judgments of ends that are stated in terms of observable pupil behaviors. They are viewed as a means of identifying goals and of describing the outcomes of instruction

[1] H. H. McAshan, *Writing Behavioral Objectives: A New Approach* (New York: Harper & Row, Publishers, 1970), p. 97.

in terms of performances that children *should* have as the result of their participation in an instructional activity.

According to this view, any worthwhile effort to improve educational programs must include proper identification and description of specific behavioral objectives if the program is to be evaluated. For, unless those objectives unique to a given area of study are clearly stated, the pupil, the teacher, and the program planner will not know what is to be done, how it is to be done, or when the goal has been accomplished.[2] Behavioral objectives are seen, in short, as a significant way of helping everyone who is involved in an educational activity to clarify what the goals of that activity are and what criteria to employ in determining if those goals are attained. The proponents of behavioral objectives tacitly assume that *instructional goals can be stated in terms that command agreement as to their reference by all who use them.*

Major Sources Giving Rise to Behavioral Objectives

It is significant to point out that the impetus for behavioral objectives has emerged from *outside* the particular subject fields to which they are generally applied. In the main, the behavioral objectives movement represents an amalgam of systems analysts, behavioral psychologists, and measurement theorists who, from their respective vantage points, require that outcomes of an activity be measurable. In turn, these forces have had a significant impact upon curriculum decision-making in recent years, where the criteria of efficiency and effectiveness are yardsticks for evaluating instruction. *There are constant pressures to give priority to those educational goals that can be most effectively and efficiently attained, to do that "which works."*

Views of Knowledge and the Learner

Further embedded in the issue of behavioral objectives are substantive matters pertaining to *how* and *what* we learn—to

[2] *Ibid.*, p. 5.

the psychologies and philosophies that guide us. Behavioral objectives arise in part out of, and are clearly nourished by, stimulus-response connectionist learning theory in which learning is viewed mechanistically and where precise statements of fact, careful observation, and limited induction are accepted as paradigms of knowledge and thus as paradigms of human behavior. This framework, of course, contrasts with those views of learning which accommodate the imaginative, the intuitive, and "discovery" as fundamental attributes in the construction of knowledge.

Accordingly, from the behaviorist's view, knowledge is presumed to be certain and absolute as opposed to uncertain and relative; it is also presumed that knowledge is impersonal (the same for all) as opposed to the view that experience is transformed into personal schema which further guide an individual's interaction with the world about him. *From the behavioral viewpoint the child is regarded largely as the recipient of knowledge, rather than being seen as an active participant in its creation.*

A Brief Critique of Behavioral Objectives

I have briefly reviewed some underlying aspects of behavioral objectives which lend perspective to a more practical examination of them. Three underlying features of behavioral objectives that have been pinpointed are (1) there is an assumed consensus about the goals of instruction; (2) an emphasis is placed on those instructional activities that "work" and are measurably workable; and (3) what is learned and how it is learned are regarded essentially as the same for all.

The consequences of these points in respect to curriculum development and teaching method raise a number of significant questions. First, from the behavioral objectives viewpoint, it can be argued that only that which can be tested is that which can be stated objectively; the inference is that only what can be tested shall be taught. Given this argument, the curriculum developer's task and the teacher's responsibilities are clearly delineated, if not prescribed. When this mandate is applied in extreme form, a premium is put on the teaching of information,

on rote recall of definitions, rules, and principles, and on the solution of problems with only one solution. (How can divergent thinking outcomes be stated in behavioral terms; as, for example, "How many things can you do with a paper clip?") Learning thus becomes a "planned happening."[3] The adherence to behavioral objectives necessarily precludes the belief that learning and knowing can have tacit (personal) dimensions as well as explicit ones.[4]

Second, there is a belief that observed performance in an instructional setting is conclusive evidence that the pupil has mastered a behavior which has been taught. But, children (and adults) sometimes behave as though they have learned when they have not; while, more significantly, they sometimes learn witho overtly displaying behaviors that would give testimony to the learning. Coupled with the realization that students have different capabilities for learning and different attitudes toward learning, it seems presumptuous to apply the same educational standards upon which behavioral objectives generally are based to all children in all circumstances.

Moreover, teachers also differ in their abilities to teach and in their styles of teaching; they bring to an instructional setting a range of competencies and backgrounds which parallels that of their students. Where behavioral objectives structure the child's learning situation, they in turn constrict the teacher's range of instructional alternatives. *Both* teacher and pupil can become constrained by the planned happening.

Third, there is a belief that we either know or can readily identify the educational objectives toward which we ought to strive. Any curriculum that is used by real people will have outcomes that cannot be anticipated. Of those outcomes that can be readily identified, fewer still are readily translatable into behavioral terms. *A danger exists that a curriculum becomes only those outcomes that can be readily specified.* As Atkins has so clearly said:

[3] James B. MacDonald and Bernice J. Wolfson, "A Case Against Behavioral Objectives," *Elementary School Journal* 71 (December 1970), 119-28.

[4] Harry S. Broudy, "Can Research Escape the Dogma of Behavioral Objectives?" *School Review* 79 (November 1970), 43-56.

> . . . it is difficult to resist the assumption that those attributes which we can measure are the elements which we consider most important. . . . The behavioral analyst seems to assume for an objective to be worthwhile, we must have methods of observing progress. But worthwhile goals come first, not our methods for assessing progress toward these goals. Goals are derived from our needs and from our philosophies. They are not and should not be derived primarily from our measure. It borders on the irresponsible for those who exhort us to state objectives in behavioral terms to avoid the issue of determining worth. Inevitably there is an implication of worth behind any act of measurement. What the educational community poorly realizes at the moment is that behavioral goals may or may not be worthwhile. They are articulated from among the vast library of goals because they are stated relatively easily. . . . Let's not assume that what we can presently measure necessarily represents our most important activity.[5]

Behavioral Objectives and Reading

So far I have made a brief explication of behavioral objectives without making particular reference to reading. My reason for doing so has been to keep the focus of our attention on the matter of *behavioral objectives* as the basis of our discussion and to avoid the possible diversion of our attention to theories of reading, a significant issue in itself. In order now to put reading into the context of behavioral objectives, I will assert that reading is a two-fold process that involves the transformation of written forms of language *to* meaning—that is, the *decoding* of and the *comprehension* of the printed message.

Aspects of the decoding process are, of course, more suitable to direct observation than are aspects of the comprehension process. For, while an individual's abilities to decode may be manifested through various word recognition tasks, the observation of his abilities to comprehend is less certainly grounded. Does not the *meaning* of a work of literature, for example, become known only to the individual who interacts with it?

---

[5] J. Myron Atkins, "Behavioral Objectives in Curriculum Design: A Cautionary Note," *Science Teacher* 35 (May 1968), 27-30.

Measuring *attitudes* toward reading is an even more elusive procedure. It seems clear that—in respect to reading—decoding, comprehension, and attitudes can be cast into behavioral terms only at the cost of decreasing levels of certainty as to the reliability of the measurement as one progresses from skills to understanding to affect.

A Summation

I should like to close with some observations that I think should be considered regarding behavioral objectives *vis-a-vis* reading. First, I believe that there is an ever present tendency to regard the demonstration of reading skills which *can most easily be observed* (for the most part, the decoding skills) as the reading process. The ability to read is not simply the sum of word recognition skills; *knowing* is more than the sum of the processes which lead to it.

Second, how the child is perceived, how and what he learns, and how schooling should be structured, as viewed through the template of behavioral objectives, warrant careful consideration. A rigid adherence to behavioral objectives, if not prescribing one's view of reading, at least puts blinders on the viewer and narrows his range of visible alternatives.

Third, there is a growing fetish among proponents of behavioral objectives to regard the *process* of defining behavioral objectives as an end in itself and to assume that there are simple procedural solutions to the many complex problems of education. Behavioral objectives, of themselves, are neither good nor bad, but can be either, by those who use them.

For reading in particular, behavioral objectives ought to be viewed in terms of their limited utility, particularly in respect to those *skills* where they have the most utility. As to the higher cognitive functions involved in reading comprehension, as well as in the affective aspects of reading, their limitations should be recognized.

Reading researchers, developmental psychologists, and linguists are now exploring new frontiers of knowledge about learning, language, and reading and raising significant points about the

efficacy of the assumptions and theory underlying behavioral objectives. Perhaps one of the principal but unforeseen outcomes of the behavioral objectives movement is that it causes us to ask ourselves what we *truly* believe about the nature of the child and the function of schooling.

REFERENCES

Aerasian, Peter W. "Behavioral Objectives and the Teaching of English." *English Journal* 60 (April 1971), 495-99.

Guth, Hans P. "The Monkey on the Bicycle: Behavioral Objectives and the Teaching of English." *English Journal* 59 (September 1970), 785-92.

Hersh, Richard H., and Cohen, Stuart J. "A Case Against a Case Against Behavioral Objectives." *Elementary School Journal* 71 (May 1971), 430-37.

Krathwohl, David R. "Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Materials Development." *Journal of Teacher Education* 16 (March 1965), 83-92.

Maxwell, John, and Tovatt, Anthony (eds.). *On Writing Behavioral Objectives for English*. Urbana, Ill.: National Council of Teachers of English, 1970.

Purves, Alan C. "Of Behaviors, Objectives, and English." *English Journal* 59 (September 1970), 793-97.

TESTING IN READING:

A GENERAL CRITIQUE

Never in the history of education have reading tests enjoyed as much status as they do today. They provide a data base used increasingly as a means (often the sole means) of evaluating pupil progress, teacher effectiveness, and program success. They are used in research studies to compare methods and materials. They are linked by law in several states with special or basic state support for the schools. Schools and school systems are publicly compared on the basis of rankings of pupil populations on reading tests. Election campaigns often center on pupil performance on reading tests. Publishers and private contractors are sometimes paid on the basis of student performance on reading tests.

It is always desirable to re-evaluate the evaluators we use from time to time. With so many crucial educational decisions being based on reading tests, this re-evaluation becomes urgent.

Uses of Reading Tests

Above, reference has been made to some of the current uses of reading tests. Only two basic uses of reading tests are legitimate. They are as follows:

(1) To measure the effectiveness with which any person uses

reading to comprehend written language. Within this, the two main concerns are (a) flexibility in comprehending a wide range of materials and (b) degree of proficiency as compared to other readers or as compared to some absolute scale of proficiency in comprehending written language.

(2) To diagnose the strengths and weaknesses of readers as an aid to planning instruction which will help to make them more effective.

Testing for each purpose will vary depending on the theory of the reading process, and of reading acquisition which the tester uses. In some cases, readiness tests will be used if the tester believes that there are nonreading tasks which must be mastered as prerequisites to successful acquisition of reading.

A major weakness of current reading tests is a failure to articulate views of the reading process and of learning to read as a basis for building the tests, subtests, and test items.

Tests are often built on eclectic traditions of what is important in reading and learning to read. These are sometimes derived directly from popular instructional reading programs; but just as often the instructional programs derive their rationale from the tests on the theory that if something is commonly tested it must be important. This misuse of tests results in a self-justifying cycle which institutionalizes tradition.

That cycle tends to block progress in improvement of reading instruction through the application of new insights from research, theory, and practice. The tester says we must test what is being taught; and the teacher says we must teach what is being tested. Innovative programs are judged on the basis of performance by pupils on traditional tests which incorporate the same faults that the new programs seek to overcome in old programs.

Since tests grow from tradition rather than articulated theory, they develop subtests with large areas of overlap, while leaving gaping holes that are not tapped at all.

The successful reader is treated as a possessor of bundles of skills rather than as a user of written language. Traditional semilogical, sequencing criteria and hierarchical arrangements are imposed on these skills which are isolated, for ease in testing, out of any context of language use which they may have.

In the absence of a strong base in reading theory, current

reading tests substitute sophisticated test theory. Surrounded by norms, percentiles, measures of significance and other statistical armor, the tests give an impression of scientific validity which conceals their hollow cores.

Tests, any tests, will produce statistical results with populations that take them. These results can be mathematically manipulated. By adjusting the test items on the basis of the statistics they produce, one may achieve neater statistical patterns. But in fact one may never draw conclusions whose significance go beyond the validity of the assumptions on which the test is based.

Criterion-referenced tests, those which measure achievement of stated goals rather than comparing to a statistical norm, are even more in need of being rooted in a strong theory. In reality, they tend to be selected skills arbitrarily sequenced.

Statistical Fallacies in Reading Testing

There are a number of key statistical fallacies that are widely incorporated into justifications for misuses of reading tests. A few will be explored here.

Norming over diverse populations. Sophisticated test theory dictates that norms or percentiles should be developed by administering the test to a structured sample of the general school population. Care is taken to include the right proportion of urban, suburban and rural pupils, white, black and other; east, west, north and south, and so on. These national norms or percentiles are then published. The implication is that the test is valid for use nationally. Though test-makers often suggest that schools may wish to use regional or local norms, there is a clear implication that individual pupils, classes, schools and districts may be usefully compared to the national norms.

But now let us introduce just one condition. Suppose that the tasks and questions on the tests are selected so that they favor one group (white, suburban, middle-class eastern pupils) over all others. This could be the result of choosing to deal with experiences and concerns more common among the favored group. Differences between groups then would be at least partly the result of the relative relevance of the test and not any actual difference in reading effectiveness. Furthermore, simply using local

norms would not remedy this problem, since pupils' scores would reflect the degree to which they matched the favored group in background. Now add to this problem regional and social dialect differences, and the justification for national use of the test through a structured sample is even more highly suspect. Only if the test-maker argues that in fact all pupils should be judged by the degree to which they compare to a high status group is such norming justifiable, and such a judgment involves value questions which cannot be answered statistically.

The importance of small variations in test performance. The statistical treatment of raw scores on reading tests makes it possible to equate them to grade placement—the grade of the average pupil who achieved the score in the norming trials.

Since the test must use a limited number of items because of time considerations, the differences, particularly at the upper and lower ends of the scale between the average score in two adjacent grades, may be only a few items. One more question right can add several months to the grade equivalence of a pupil. Consider this in relation to the relevancy questions raised above and it is clear that a slight bias against a group can explain statistically significant differences in group means.

Sky-hooks, split-halves, etc. To cope with questions of whether tests are testing what they should and doing it consistently, a number of statistical devices have been employed. One used particularly for new tests is to correlate them with other older tests. If a high correlation is achieved, then validity is assumed. However, if the new test is in fact measuring what the old test did, then why is a new test needed? And if the new test employs new insights, why expect it to correlate with the old? This sky-hook method of anchoring tests to each other clearly says nothing about the extent to which reading is really being tested. A current federally funded project seeks to establish a new test to which all current tests could be correlated. Such a test, appropriately called an anchor test, would surely anchor reading permanently to the past. Similarly, using split-half techniques to prove that a test is consistent within itself proves only a symmetry on whatever biases are built into the test and does not offer evidence about the value of the test.

Hard Is Hard But Why Is It Hard?

Test theory requires that some items should be missed by most pupils, and some items by a few, with the rest of the items ranging in-between. Further, the high scorers should be the ones getting the hard items right, and the low scorers should be the ones getting the easy items wrong. Close examination of reading-test items reveals that the items are often difficult for irrelevant reasons: ambiguity, equally correct wrong alternatives, and so forth. The fact that few people get them right may indicate that they are hard, but it may also indicate that they are irrelevant or poorly written. The fact that the right pupils get them right may demonstrate more that high scorers are good at thinking like test writers than that they are better readers. Again, what is important is that statistical evidence cannot substitute for intrinsic criteria in judging the relevance or difficulty of items.

Related to this statistical fallacy is an artifact that results from weighting certain items by virtual repetition (a series of very similar items). A pupil tends to get all like items right or wrong if in reality his performance reflects knowledge or lack of knowledge. A minor lack becomes magnified into a major weakness. An example of this is the syllabication sections of certain tests.

Averaging Ends and Means

A statistical fallacy occurs in many reading achievement tests when an overall score is calculated which combines scores on "skill" subtests with those on comprehension. Since skills are ostensibly the means by which comprehension (the end product of reading) is achieved, such a score is meaningless.

Counting in Diagnosis

Statistics which produce summary scores are much easier to manipulate than those which relate to complex phenomena in detail. The effective use of diagnostic testing is often defeated by being more concerned with quantity of errors or a grade level equivalence, than with the specific phenomena revealed by per-

formance on the reading tasks involved. Number of reading errors or items wrong gives little diagnostic information.

Design Problems in Constructing Reading Tests

Aside from the statistical fallacies confusing test use and misuse, there are a number of design problems which test-makers have not yet solved adequately.

Convergence. One of the most difficult design problems in writing tests is that there must always be a right answer. This leads to focus on convergent responses—on those which match the preconceptions of the test-maker. Two groups are hampered by such tests: culturally divergent groups whose experience does not match the testers, and creative thinkers who are able to see "different" relationships. If you march to the beat of a different drummer, the test penalizes you.

Pupils who know too much. Multiple choice responses are designed to mislead pupils with common misconceptions. Since a misconception is better than no concept at all, pupils are penalized for knowing a little and will be wrong more than chance would predict. Even worse off is the pupil who knows more than the test-maker. He will often reject the "right" answer because he recognizes it as a misconception or oversimplification.

What they learn vs What they know. In testing comprehension, it is easy to end up testing general knowledge. The pupil may be able to answer the questions without reading the test selection. To overcome this problem requires a measure of prior knowledge or tests on material all pupils lack background for. The latter is a virtually impossible task.

How test-wise are the subjects? Pupils vary greatly in their control of devices for scoring higher on tests. Only some pupils have learned simple devices like skipping troublesome items, quick identification of tasks, eliminating obvious wrong choices to narrow the range of possible options, going to questions without reading test paragraphs or answers without reading questions, and so forth. There appears to be no way to neutralize fully this effect, which is also linked with the pupil's basic desire to do well (or his indifference).

Honesty. Related to the latter problem is one of honesty.

This is a complex problem because many pupils avoid using techniques which would produce higher scores. They think that is like cheating. In fact, those who give reading tests often behave hypocritically. Test-makers discourage guessing, tell pupils to read each item before looking at the answers, refer to the tests with smaller children as "a game we're going to play." They tell pupils that the test is only to help them and that the score they make is unimportant.

But in fact the score is the only part of the test in which the test-makers are interested, and decisions are made on the basis of the test scores, which may well effect the learner adversely—placement in a low track, for example. The pupil who is honest and trusting becomes a statistical victim.

How they do vs What they do. Since test-making involves counting right answers, there is a tendency in trying out test items to ignore the basis on which pupils respond to questions. In many cases, pupils are producing both wrong and right answers for the wrong reasons. Subtests turn out to be testing something quite different from what they claim to test. Auditory discrimination tests, for example, turn out to be testing largely the ability to deal with abstractions. Some pupils on such tests will resort to matching spelling patterns, producing a fair number of right answers without being able to abstract sounds from sound sequences.

Making the test clear. Pupils frequently do not understand what the task is that a particular subtest requires of them. It is quite likely that this accounts for a considerable amount of the variation of performance, particularly among younger children.

Distortion of tasks. Finding a format for test items which is suitable for inclusion in a group-administered reading test frequently results in a distortion of actual reading tasks. Some examples follow:

(a) *Items too short.* Research on reading miscues has demonstrated that short items are harder to read than longer ones because reading involves building up expectations on the basis of redundancies.[1] A sentence is proportionately harder to read

[1] D. Menosky, "A Psycholinguistic Description of Oral Reading Miscues Generated during the Reading of Varying Portions of Text by Selected Readers from Grades Two, Four, Six and Eight" (Doctoral dissertation, Wayne State University, 1971).

than a paragraph, a paragraph harder than a page, and an isolated word hardest of all. Since short items predominate in tests (words, phrases, sentences), reading test items will be harder than reading stories or other natural materials.

(b) *Words in isolation* are particularly hard to "read" because there are no grammatical cues from the sentence structure or meaning cues from the context to help identify the word meaning, yet many subtests deal with isolated words.

(c) *Comp chension questions that can be treated like nonsense.* Many questions are stated in such a way that readers may answer them by transforming the question to a statement and searching the text for a match without necessarily understanding. They manipulate the sentence patterns as if they were nonsense like the jabberwock. Q—What did the momeraths do? A—The momeraths outgrabe.

Abuses of Tests

There are uses of tests which clearly violate the publisher's advice on limitations in their use.

One common, recent abuse is giving tests at too frequent intervals. Many requirements for reporting progress in reading as often as once a month are built into contracts for research or demonstration projects. In the name of accountability, tests are being used to measure small increments of progress which they are simply not designed to handle. Factors such as regression toward the mean (a tendency for high and low scores to move toward the mean on repeated testing) become very important. Immediate, often temporary, results become more highly valued than long-term, permanently held gains.

If jobs, funding, and professional status and pay are made contingent on pupil performance on reading tests, then the tendency to teach to the test and to build curricula around the test will become a major trend. Instead of the curricular goals being centered around effective reading, the goals become performance on specific tests. Instead of tests functioning as a measure of achievement, they are turned into ends in themselves. Even if they had a sound theoretical base, that would be unfortunate. In their current state, it could be tragic. It could lead to a new kind of widespread functional illiteracy.

Another abuse of tests which we have touched on earlier is the use of test scores without close examination of each pupil's test performance. It is not enough to say Mary Lopez is reading on the 2.2 level. Her responses to subtests and items must be examined closely so that her strengths and weaknesses are revealed. Standardized reading tests are given wholesale to masses of students. But their results must be interpreted for each learner if they are to be useful in improving that child's reading. Every child has a right to be treated as an individual and not as a test statistic.

A related problem is the use of tests as exclusive means of evaluating pupils' reading effectiveness, ignoring more extensive evidence of competence because it is less easily quantifiable. Teachers will frequently treat a low scorer on a standardized test as a poor reader, even though they observe him functioning as an adequate reader every day in class. The quantifiable performance on tests is so intimidating to the teacher that he will not trust his own professional judgment.

Often the pupils perform poorly on the test because it is irrelevant to them and penalizes them for linguistic, experiential, and cultural differences (not deficiencies). Instead of rejecting the test as irrelevant, wholly or partly, teachers and administrators accept the test and misjudge the achievement, strengths and weaknesses of the pupils. Programs are then planned to remediate deficiencies that never really existed.

Other abuses of reading tests occur in evaluating new methods and instructional materials. Frequently, little consideration is given to the basic soundness of the method or the materials or the principles on which the tests are based. Rather they are judged largely by how well pupils do on pre- and posttests. While effective instruction must ultimately be judged by the learning it produces, progress in improving instruction cannot come by using a trial-and-error technique for evaluation Not all programs are worth trying, nor can the test results be usefully interpreted if the instructional program is not thoroughly analyzed.

If instructional methods and materials are built around tests, it is likely that pupils will improve their performance on the tests. The most extreme version of this is to use the test items as the instructional program, asking the pupils to respond over and over until they produce right answers all or almost all the time.

This ability to do well on the test is then assumed to prove learning has taken place because the learner now can produce a test score characteristic of a proficient reader. Performance is assumed to reflect the reader's competence exactly, no matter how it is produced.

What is not understood is that all behavior is the end product of a process and that competence is not behavior, but control over the process. Behavior, in the form of test performance, can be used to infer what competence exists in reading, but this requires an interpretation of the behavior based on understanding how reading works.

Deciding, on the basis of unexamined reading-test scores, such vital aspects of the child's future as the class, group, or track to which he will be assigned is a terrible abuse of reading tests. It jeopardizes the pupil's future and does not even offer a basis for improving his reading proficiency, since pupils who are very different in reading may achieve similar scores.

Reading Theory: Key Questions Test-Makers Must Ask

The earlier parts of this paper have portrayed reading tests as rather primitive, eclectic, and atheoretical in all aspects except for their use of sophisticated test theory.

The questions reading test-makers must deal with to produce better and more useful tests are clear; however, there is no agreement yet on the answers.

Major questions that must be answered in building better tests are (1) what is reading; (2) what are the essential skills and strategies that a successful reader must possess; and (3) what are the purposes and uses of reading? These will be considered in order.

What is Reading?

Elsewhere the author has stated that "Reading is a complex process by which a reader reconstructs, to some degree, a message encoded by a writer in graphic language." [2]

---

[2] Kenneth S. Goodman and Olive S. Niles, *Reading Process and Program* (Urbana, Ill.: National Council of Teachers of English, 1970), p. 5.

Whether one accepts this definition, which carries with it the concept that reading must result in meaning to be considered reading, or some other definition, one must still base test construction on some coherent definition.

The following group of related questions must also be answered:

(1) Can reading skill (for example, matching letters to sounds) be separated from the quest for meaning in teaching or testing?
(2) At what point can reading as a process be separated from its uses?
(3) Does the reading process necessarily involve oral language at all, or is it entirely a matter of deriving meaning from written language?
(4) Is the reading process different at various stages of development, or is it the same, varying mainly on the effectiveness of the reader?
(5) Is reading a general ability, or is it one which varies with content, interest, or task within each reader depending on his own background?
(6) Is the reading process the same or different across people, languages, cultures, or orthographies?

Though there are implicit answers to these questions in many current reading tests, it appears that the test-makers have made assumptions often without considering the issues involved.

What Are the Essential Skills and Strategies that Effective Readers Possess?

Reading tests have generally employed subtests to get at what are assumed to be essential reading skills and to monitor their development. To justify such practice, the following questions must be answered:

(1) Can essential skills or strategies be isolated for testing without changing their relative values, their basic uses, or the reading tasks in which they occur?
(2) Are such strategies or skills universal across people, contexts, purposes, languages, and orthographies?
(3) Is there an essential sequence in learning to read; i.e., must some skills or strategies be learned before others?
(4) How are reading skills or strategies to be understood in terms of how language works and is used?

What Are the Purposes of Reading?

Language, including reading, is always a means and never an end in itself. This is true whether one is talking about the proficient user or one just learning. Meaning, either its expression or comprehension, is always the end for which language is the means.

Ultimately then, any reading test must measure the success of the reader in comprehending written language. It is meaningless to consider performance on skills tests a measure of reading achievement. What counts at all stages of development is what the reader understands as a result of reading.

Test-makers must be concerned with the following questions:

(1) What is comprehension; how does it work, how is it achieved, how varied is it?

(2) What different problems face the reader who is reading to acquire knowledge, as compared to one who is reading for a message already within his conceptual grasp?

(3) What role does the reader's background and interest play in successful reading?

(4) How does critical reading differ from other reading?

Future Reading Tests

Diagnostic reading tests in the future will need to focus on reading as it really occurs in natural language. This suggests the type of task now found in informal reading inventories. But the diagnostic test of the future will be designed so that the strengths as well as the weaknesses of learners will be made clear. A shift will need to be made away from counting errors to analysis of performance, to get at the underlying competence.

Achievement tests will need to deal with comprehension in a range of reading situations. They will need to avoid irrelevance. And they will need to get at the reader's ability to use written language effectively. Group tests may well disappear. They sacrifice too much for the sake of economy of time.

Still, however, the main improvement needed in the area of testing is in use. No test, however cleverly it is constructed, can substitute for the insights professional teachers get from working closely with children.

PERFORMANCE CONTRACTING:
A ONCE-OVER

Reactions to the performance contract made between the Texarkana schools and the Dorsett Educational Systems in 1969 varied nationally from charges that such contracts would dehumanize education to countercharges that education was too important to be left to educators. Whereas school contracts routinely provide for services and equipment, easily auditable, the Texarkana-Dorsett contract called for the delivery of increased performance on the part of school children. The Texarkana-Dorsett product was to be a group of better readers, a matter far more difficult to audit.

This first performance contract seemed to threaten the domain previously reserved exclusively for classroom teachers. Traditionally, the teacher has been considered to be the one most knowledgeable about what the child needs. It is unfortunate that all too many teachers are willing to accept a large number of failures as a normal consequence of well-conducted classroom procedure. School board members, on the other hand, have become "less confident that teachers are dedicated to teaching students . . . increasingly aware that they, perhaps even more than any other group in public education, are being held accountable for what transpires in schools."[1] School superintendents in major

[1] "School Board Members View Performance Contracting," *Arizona Teacher* (January 1971), 18.

cities, such as Nolan Estes of Dallas, have begun stating that schools are going to have to deliver one year of educational growth for one year of instruction, no matter what it takes.[2] Former Commissioner of Education James Allen involved the federal government in another human right: the right of every child to learn to read to the full extent of his potential. Mounting need, increased public concern, conceded failure of the present system to produce desired results—all made a new approach imperative, and performance contracting entered the educational scene.

Critics of the initial Texarkana Project felt vindicated when results of the 1969-70 contract were announced. Though the $105,000 paid to the Dorsett Corporation produced some results, these were declared by many to be invalid. Dr. Robert Kraner of the EPIC Evaluation Center, Tucson, the final evaluator, "concluded in his final report that 30 to 100 per cent of questions in the tests administered to Texarkana students in May, 1970, were 'contaminated,' meaning that they had been taught in the classroom prior to testing. Said Kraner: 'The teaching of test items, or closely related test items, has invalidated the test results to the extent they cannot be used as a valid measure of achievement.' Loyd Dorsett replied that only 'a small amount of improper teaching' went on, which meant that only 7 per cent of the test questions could be considered invalid."[3]

Such criticism did not shake Texarkana's faith in the principle of performance contracting. Dorsett bid for a second year's contract and was refused. "There are rumors floating that Texarkana may go to court to recover $110,000 already paid to Dorsett."[4] A contract was awarded, however, for the 1970-71 school term by Texarkana to Educational Development Laboratories, Inc.

Very simply stated, a school contracts with industry to assume responsibility for instruction in a given area, and industry guarantees increased performance on the part of the students involved. Payment to the contractor depends largely on the

[2] Triple-T Conference Held at Southeastern State College, Durant, Oklahoma, September 1970.

[3] Ramona Weeks, "Performance Contracting: Pitfall or Panacea?" *Arizona Teacher* (January 1971), 18.

[4] *Ibid.*, p. 19.

amount of success he has, as revealed by objective measurement of student performance. Contract patterns vary from the establishment of special learning centers conducted by employees of the contracting firm to the utilization of regular staffs and facilities under the direction of contracted managers. All kinds of funds—local, state, and federal—are being used to pay for the contracts. All existing contracts provide for a turnkey operation at the end of one or two years usually. Performance contracts, then, are used to get new means of guaranteeing educational success into the political system.

Undeterred by the questionable results in Texarkana, a significant number of schools all over the country entered into performance contracting for the 1970-71 school year. A list of the schools, contractors, and programs is given below. Of interest to reading specialists is the significant fact that every program on the list has a reading component. It would seem imperative, then, that reading teachers be well aware of the nature of performance contracting.

Far from the simple contract for services or products which most attorneys can execute for school districts, performance contracts are exceedingly complex. Several management firms are now engaged in the preparation of contracts and the supervision of the job to be done. Jack Stenner, director of the Management Support Services, listed the advantages of performance contracting as follows:

1. It facilitates the targeting and evaluation of educational programs. . . . It fosters the objective evaluation of educational results and also the managerial processes by which these results were achieved.
2. Performance contracting for instructional service could introduce greater resources and variability into the public school sector.
3. The performance contract approach allows a school system to experiment in a responsible manner with low costs and low political and social risks.
4. The right of every child to read at his grade level will undoubtedly place great burdens upon the schools' limited resources. If the Nation's schools are to make this principle a reality, they might want to consider using performance con-

OPERATIONAL PROGRAMS FALL 1970/STUDENT ACHIEVEMENT[5]

Educational Agency	*Learning System Contractor*	*Program*	*Students*		*Target*
			Number	*Grades*	*Payment*
Boston (Roxbury), Mass.	Educational Solutions	Reading	400	K-6	$ 80,000
Colorado, State of	Dorsett Educational Systems	Reading	300	6-8	50,000
Cherry Creek	Dorsett Educational Systems	Reading	100	6-8	...
Denver	Dorsett Educational Systems	Reading	100	6-8	...
Englewood	Dorsett Educational Systems	Reading	100	6-8	...
Dallas, Tex.	New Century	Reading & Math	875	9-12	...
Dallas, Tex.	Thiokol	Occ. Skills & Motiv.	875	9-12	...
Flint, Mich.	Dealer for E.D.L. Materials	Reading	2,160	9	210,000
Gary, Ind.	Behavioral Research Laboratories	All subjects	800	K-6	640,000
Gilroy, Calif.	Westinghouse Learning	Reading & Math	103	2-4	60,000
Grand Rapids, Mich.	Westinghouse Learning	Reading & Math	400	1-6	143,700
Grand Rapids, Mich.	COMES	Reading & Math	600	6-9	164,000
Greenville, S.C.	COMES	Reading	480	6-9	100,000
Jacksonville, Fla.	Learning Research Associates	Reading, Math, Social Studies & Science	300	1	70,000
Oakland, Calif.	Education Solutions	Reading	400	6-8	80,000
Philadelphia, Pa.	Behavioral Research Laboratories	Reading	20,000	1-2, 7-8	800,000
Providence, R.I.	New Century/Communications Patterns	Reading	1,500	2-8	145,000
Savannah, Ga.	Learning Foundations	Reading	1,000		97,000
Texarkana, USA	Educational Developmental Laboratories	Reading, Math & Dropouts	300	7-12	100,000
Virginia, State of	Learning Research Associates	Reading & Math	2,500	1-9	212,500
Norfolk	Learning Research Associates	Reading & Math	500	4-9	...
Buchanan Co.	Learning Research Associates	Reading & Math	500	1-7	...
Dickinson Co.	Learning Research Associates	Reading & Math	250	1-7	...
Lunenberg Co.	Learning Research Associates	Reading & Math	250	4-7	...
Mechlenburg Co.	Learning Research Associates	Reading & Math	250	4-6	...
Prince Edward Co.	Learning Research Associates	Reading & Math	250	4-6	...
Wise Co.	Learning Research Associates	Reading & Math	500	4-9	...

[5] George R. Hall and James P. Stucker, "The Rand/HEW Study of Performance Contracting," *Compact* (February 1971), 8.

tracting for the development and validation of new reading programs.

5. Through the use of the performance contract approach, many of the previously segregated children will have their academic deficiencies removed on a guaranteed achievement basis while they are attending the newly-integrated schools.
6. The approach creates dynamic tension and responsible institutional change within the public school system through competition. Boards of education can finally establish policy and choose among alternative instructional programs.[6]

Whether the charges that performance contracting will dehumanize education and is an actual threat to classroom teachers are valid or not is undecided. Certain disadvantages, however, do seem to be apparent. There is the ever-present danger that in some way teachers or contractors will find a way to teach to the test. Since Texarkana, most contracts include the services of an independent evaluator and contain intricate provisions to assure that the nature of the test to be given remains unknown to the contractor and the instructional staff until it is administered. Dallas, for instance, has awarded its evaluation contract to Educational Testing Service, but plans to bring in two additional evaluators to check the validity of the results.

Another disadvantage, as yet unresolved, is the matter of who holds copyrights on materials produced or reports published about the project.

A third disadvantage, probably not as critical as the others, is the matter of responsibility. To whom do citizens complain in the event of problems or dissatisfaction? It would seem that the chief schools officer of the contracting school would still bear the responsibility, but it is logical to assume that contracted arrangements would allow for a shift in responsibility.

Fourth, without guidance from a profession of reading specialists, it is possible that a school and a contracting firm may arrive at some unsound behavioral objectives, and though the objectives stated in the contract are met, the end results may be disappointing if not disastrous.

---

[6] Jack Stenner, "Performance Contracting for Instruction" (unpublished manuscript, 1971).

Fifth, the charge has frequently been made that performance contracting is yet another way of lining industry's pockets with money at the expense of schools, teachers, and children.

Sixth, there is the possibility that Robert Stake and James Wardrop are correct in feeling that progress made may not be entirely due to the efforts of the contracting firm.

> Suppose that three students were to be tested with a parallel form immediately after the pretest. The chances are better than 50:50 that at least one of the three would have gained a year or more and appear ready to graduate from the program.
>
> Suppose that 100 students were admitted to contract instruction and pretested. After a period of time involving no training, they were tested again and the students "gaining" a year were graduated. After another period of time without training, another test and another graduation occur. After the fourth such "terminal" testing—even though no instruction had occurred—the chances are better than 50:50 that two-thirds of the students would have graduated.[7]

Finally, standardized testing used in many of the contracting situations "does not have the necessary content validity for individual student assessment. For years test authors and test publishers have cautioned users against using these tests as diagnostic instruments. Performance-contract criterion tests should, in effect, be diagnostic tests." [8]

Literature concerning performance contracting makes a number of recommendations to schools who intend to enter into such a program. Because of the complexity and the newness of such programs, it is recommended that schools should employ the services of management consultants in both the preparation of the contract and the conduct of the program. Also because of the novelty of the idea, schools are cautioned to regulate the flow of visitors to the project; in many instances, numbers of interested onlookers have seriously limited the success of the program.

It is further recommended that the teacher's role in the program should be clearly defined before instruction begins.

---

[7] Robert E. Stake and James L. Wardrop, "Gain Score Errors in Performance Contracting" (unpublished manuscript).

[8] *Ibid.*

> If, in a teacher's judgment, other materials and activities are dictated, and the teacher can support the position, the agreement of the project director and the contractor is negotiable. No teacher has the right, however, to unilaterally intervene with alien procedures and revise a system which is legally agreed to and by which a contractor will receive reimbursement. Since, in most cases, the teachers involved in this kind of operation will be volunteers, making this known initially, as a condition of participation, should require no more than a signed agreement by each teacher assenting to this requirement of the contract." [9]

The literature suggests that the school and contractor would do well to involve as large a segment of the community as possible in the project. Not only will such involvement provide much-needed public support, it will be invaluable when the project is turned over to the school district to run.

Finally, the school should engineer lower cost factors by providing a number of students which would make for the most economical situation. It is, further, uneconomical for a school to purchase any nonconsumable materials during the first year of operation; such materials should be provided by the contracting firm or leased.

A number of legal considerations enter into any program of performance contracting. Not the least of these is the relationship of the contract to local and state education agencies. Before the contract in Dallas could be approved by the Texas Education Agency, for instance, a ruling by the state Attorney General had to be made. The ruling was that contracts could be made and paid for with state and federal funds; however, all members of the contract instructional staff had to be certified teachers in the State of Texas, the program had to come under the regulations of the Minimum Foundation Program and was subject to Agency accreditation rulings. It is obvious that contracts will differ widely around the country because of varying opinions about their legality.

The management consultant, Jack Stenner, posed the following legal questions as pertinent to the preparation of the Dallas contracts:

---

[9] Albert V. Mayrhofer, "Factors to Consider in Preparing Performance Contracts for Instruction," *Educational Technology* (January 1971), 48.

1. If an outside contractor is utilized by the school district for the provision of instruction and/or services for pupils, will the school district's statutory obligation for said instruction and services be fully met?
2. Are there state or federal limitations regarding the use of State, Local, and Federal funds to support all or part of a performance contracting project?
3. Do existing state plans for various federal programs, that may be potential funding sources for the project, permit utilization of outside contractors who will provide all or part of the instruction and/or services included in the project?
4. Will the State Education Agency authorize credit for instruction provided by the contractor in reading or mathematics?
5. Will the State Education Agency permit paraprofessional staff members employed by the contractor to perform certain teaching tasks under the direct supervision of a certified teacher?
6. Will the State Education Agency permit the teacher in the contractor's program to be the contractor's employee?
7. Will an outside contractor's instructional personnel stand *in loco parentis* to the student?
8. Are the contractor employed teachers required to be certified?
9. Does a local district have authority to contract with or enter into agreements with an agency of the federal government?
10. Does a local district have authority to contract with or to enter into agreement with private concerns for the provision of instruction?
11. Does a local district have authority to expend local maintenance funds to pay for part or all of the cost of instruction provided by a private concern under contract with the district? [10]

Three types of negotiations are available to schools in the preparation of performance contracts. First, the school may decide on one single source of material and method. The advantage of being able to negotiate rapidly with a single company may be offset by the disadvantage of limiting the scope of the project to what that particular company can do. A second type of negotia-

[10] Jack Stenner, "Education Performance Contracting" (unpublished manuscript, January 1970).

tion is referred to as the modified sole-source method; two or more companies are asked to submit bids, and one of the group is chosen, or several are asked to work in concert with each other. The third type, the competitive system, requires far more time and is complicated, but it does insure the school the opportunity of investigating many sources of help and choosing the best alternative. In such a system, criteria for the program would be decided upon and published by the schools. Bids would then be taken from any number of competitive companies. Such competition is likely to produce some economy in the program also.

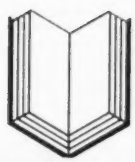
Needs and Recommendations

Because of the importance and frequency of reading programs in performance contracting, the following needs and recommendations have been formulated to provide schools some guidance in their choice of programs.

A list of sound performance objectives related to reading should be prepared. There is no indication in the literature that any professional reading group has done this. There is great need for the development of objectives that are educationally sound and encompass the issues in the teaching of reading. I have read far too many proposals of reading programs which miss the mark to want to leave this to chance. Perhaps a list of alternative objectives from which schools could select those most appropriate to their situation would be helpful.

Though the literature suggests that various methods and materials may be used in concert with each other, I saw no indication that this practice was used widely. I think we must concede that no one reading program, however good, is effective with all children in a given school. Varying modalities of learning, linguistic and experiential resources, means of perception, and individual approaches to learning provide a heterogeneity which makes a combination of systems preferable to one. There is substantial need for the identification of materials which would demonstrate to schools and teachers, long accustomed to teaching with one textbook and one method, the value of working with combinations of systems.

Criteria for critiqueing reading systems and materials would be particularly useful to school districts. Special attention should be given to soundness of programs from a linguistic point of view. None of the contracts which I read indicated that the school involved was even slightly aware that linguistic variety is a fact of life and has tremendous implication for the teaching of reading. In as much as significant work has been completed in the field of dialect and language differences, this information is critical to the teaching of reading, and such understandings should be incorporated into guidelines for performance-contracted curricula.



ACCOUNTABILITY:
A SUMMARY STATEMENT

As one examines the four previous discussions, the clear relationships that exist among them are readily evident. Both costs in education and the apparent unsuitability of present educational processes for significant chunks of our population are raising questions regarding the efficiency and effectiveness of what is being done under the label "education." These concerns about education are increasing efforts to hold accountable all those who participate in the process and product of education.

Establishing performance objectives allows one to define the task, to determine if the task has been completed, and to assess the quality of the response to the task.

Accountability for meeting specific performance objectives logically leads to the establishment of performance criteria and the development of specific measurement devices.

It was inevitable that—in our free enterprise system—someone in business would apply former Defense Secretary Charles Wilson's "get-a-bigger-bang-for-the-buck" principle to education and move unilaterally to establish relationships with school to bring it off.

The previous discussions were developed as a result of several basic questions central to the topics discussed. In addition, the very treatment of the topics subsequently raised other questions that the serious participant in education must face. Several of these questions follow:

1. Is the accountability that is being urged upon us actually too impersonal to be of benefit? Are we holding *schools* responsible for the high-quality education of our youth rather than *teachers*? If so, what kind of accountability is it? Can accountability coexist with personal anonymity?
2. Can objectives be established that encompass the total educational process and product and their effect on *people*?
3. Is there more to education than performance? And can that added something be measured? Indeed, can it be taught.
4. Is education dehumanized—for teachers or for students—when it is systematized?
5. Are our present means for evaluation in education either appropriate or adequate for the task, given the need to measure both the process and product of education?
6. Is it possible for noneducation people to do a *total job* of educating students?

Before accepting or rejecting any of these factors—accountability, behavioral objectives, evaluation, performance contracting—many issues must be explored.

Responding too quickly to the apparent efficiency of performance objectives may lead to mechanistic instruction. Rejecting behavioral objectives in the name of affective concerns may mask the fact that in too many classrooms there is little concern for or clarity of objectives related to either cognitive or affective aspects of education.

Ready acceptance of the logic of accountability may allow intrusions into education by groups with questionable motives and dubious qualifications. Rejecting the concept of accountability in the name of professionalism may mask inefficiency, sloppiness, and unconcern.

Reliance on inappropriate instruments may damage children and support inappropriate teaching. Rejection of evaluation because of the inadequacy of instrumentation can mask irresponsible teaching.

In conclusion, as it considers instruction, the educational profession has a tendency to feel itself inadequate to oppose victimizing pressures. But, when it comes to salary, demands on personal time, numbers of students per class, the profession becomes quite militant. Perhaps we need to examine our priorities, take

the good that is implicit in the factors—accountability, behavioral objectives, evaluation, performance contracting—and vigorously pursue that good for the sake of our students and their futures.